Model Phaser: Architecture

Version: 0.1 17/SEP/09

Introduction

This document will describe the architecture of the *ModelPhaser* interface (template method) in Solve-o-Matic, which can be used for the phasing of structure factor amplitudes measured in an X-ray diffraction experiment. There are two main approaches (strategies) available for this: direct refinement of the given model against the newly measured amplitides, or a molecular replacement process which will determine the appropriate positions of one or more copies of the input model in the asymmetric unit.

Purpose

The majority of X-ray diffraction data is measured as a single wavelength from a native crystal, making experimental phasing in the vast majority of cases impossible. In these cases it will be necessary to provide a model to give the initial phases: for ligand binding studies this will essentially be the known protein model, so calculation of a difference map should be sufficient to obtain an image of any ligands present. In other cases where the unit cell and symmetry are not isomorphous with the model it will be necessary to perform a molecular replacement step to determine the number and orientation of the molecules in the unit cell.

Preconditions

For the difference map calculation the following preconditions are necessary:

- The model is available
- The corresponding unit cell and spacegroup are known, by default in the CRYST1 card
- The data were reduced with the correct cell constants and pointgroup applied.

In the cases where alternate origin definitions are possible the system will need to compare the indexing choice applied during data reduction to the molecule orientation and reindex if necessary.

For the molecular replacement pipeline the following preconditions apply:

- The sequence for the protein is known (for model preparation and sequence homology calculations)
- A model or models are available
- The spacegroup of the experimental data can be estimates from systematic absences (or enantiomorphs) or all possible spacegroups for the given crystal pointgroup can be tested

Given the different domains covered by these preconditions, it should be possible to determine ab initio the correct strategy to apply for a given model and data set. Therefore, these may be reached through a single interface specification.

Postconditions

A successful outcome will be a set of phases to apply to the measured structure factor amplitudes, with a refined (partial) stucture to use for navigation. For the difference map calculation the region of interest will probably be known in advance, so the user may inspect the map to make their own judgement about the active site. Reporting the residuals from the refinement may offer a useful indication of the success of the process.

For the molecular replacement pipeline the outcomes are the same, though the most likely next step is to rebuild the protein structure into the calculated map to refine and hopefully reduce model bias.

Error States

To be determined.

Process

This will consist of a factory which will decide the most appropriate route to take through the phasing process (though this may be specified manually) which will in turn supply a template method for the given approach (difference map calculation or molecular replacement.) These will in turn include a number of possible strategies (i.e. implementations of the above interface) which make use of different software to solve what is conceptually the same problem. The selection of the strategy at this stage is poorly defined.

This will build on two interfaces: difference_map and molecular_replacement, which are composed within the model_phaser. There will be multiple implementations of these interfaces (i.e. strategies) which will be available from a library *via* a factory / builder. These will in turn depend on the library of program wrappers and hence on the underlying software.

Licensing

During development this will not be distributed. When it is "finished" the resulting pipeline should be made available to CCP4 using an EDNA framework to provide the interface.